预印本服务中的质量控制方法研究

刘静羽, 张智雄, 黄金霞, 彭媛媛

（中国科学院文献情报中心，北京 100190）

摘要：质量控制是预印本服务中的重要环节。当前国际通行的预印本质量控制方法包括“上传者和作者的研究身份审核”“论文格式完整性审核”和“论文内容学术性和真实性审核”三方面；人工审核为主、机器审核为辅。中科院 ChinaXiv 预发布平台按国际通行模式建立了质量控制机制，包括论文常规审核制、领域专家评审制、论文开放评议制、问题论文撤稿制和非诚信作者名单制等。未来可在质量控制机制公开透明、充分发挥开放评议作用、加强国际交流方面做出改进。

关键词：预印本；质量控制；预印本平台；开放获取

中图分类号：G250

1 引言

预印本因其发布周期短、传播范围广等优势，成为开放科学环境下的一种重要的学术交流模式。从第一个预印本平台 arXiv 出现以来，众多科学团体以及出版社相继推出预印本服务，服务器中的论文数量迅猛增长，2015 年 arXiv 论文量已经超过 100 万，预印本受到的关注度和在学术交流中的地位不断提高。与此同时，也应注意到，一些“非学术”或是“伪科学”论文混入预印本平台；一部分上传者为了快速发布成果，伪造研究身份，或者违背学术道德在预印本平台发布虚假、不真实、抄袭的内容。这些问题导致预印本平台内容质量的下降、预印本服务提供者信誉受到损害，更重要的是在很大程度上影响了科学家对于预印本的信任和使用，阻碍了预印本的发展。

据 Google Scholar 统计，全球已有 4/5 被引用量最多的物理学和数学内容都可以预先在 arXiv 上找到[1]，但还是有相当一部分科研人员对预印本的质量心存顾虑。学术界一直将引用包括预印本在内的未经同行评议的论文看作禁忌[2]，而一些不成熟的研究结果、未经审查在预印本系统发布，使得预印本整体来看良莠不齐更阻碍了科研人员使用预印本进行交流。此外，2017 年，根特大学副教授 Matt Shawkey 在社交网络上表达了他的观点，认为预印本的内容不应和经过同行评议的论文同等对待。另一方面，科研资助机构虽然目前普遍支持预印本发展，但也对预印本质量表示担忧。如 NIH 在 2017 年 5 月的一份声明中提到，目前阶段性成果（Interim ResearchProducts）的整体质量确实不如经过同行评审的研究成果，提示研究人员对于预印本的资源进行谨慎判断，并要求发布阶段性成果的平台须具备严格、透明的关于阻挡论文抄袭、利益冲突、学术不端行为的政策[3]。同年，美国 ASAPbio 基金会在生物学预印本中心服务库（Central Service, CS）规划中提出，该库收录的预印本必须符合基本的学术标准，包括作者署名、学术道德规则和法律、社会标准，必须正确标明基金支持等。[4]因此，只有做好预印本服务中的质量控制，才能打消科研人员对

于预印本的疑虑，才能真正促进预印本的发展。

目前全球主要的预印本平台普遍认识到质量控制的重要性，纷纷采取措施阻挡伪科学或防止错误信息的传播，本文通过调研从中总结归纳出当前国际通行的、预印本服务质量控制方法；并以此为基础，介绍和分析了我国中国科学院科技论文预发布平台 ChinaXiv 的质量控制机制，提出未来可以改进的方向。

2 国际通行的预印本服务质量控制方法

为深入了解当前国际预印本服务的质量控制方法，本研究选取 arXiv、bioRxiv、SSRN、ChemRxiv、PeerJPeprint、MDPI Preprints 六家国际知名预印本服务平台，对其网站政策与质述控制相关内容，以及注册、上传论文流程进行调研。

通过调研了解到，目前国际主流的预印本服务对平台发布的内容进行简单的质量控制，以保证论文达到基本的学术要求，包括作者具有研究身份、不造假、不抄袭；内容真实、研究过程清晰；正确标注署名等；[5]避免伪科学、虚假等违背学术道德，或学科领域不适合的论文在平台发布。质量控制措施主要包括“上传者和作者的研究身份审核”“论文格式完整性审核”和“论文内容学术性和真实性审核”三方面。操作层面，目前各大预印本服务的质量控制以人工审核为主，部分平台采用机器辅助的方式。人工审核主要负责内容学术性、真实性的审核，一般由领域专家承担；机器审核，主要是在上传者注册身份识别论文格式审核等方面发挥作用。例如，arXiv 对每一篇上传的论文进行监控，利用 arXiv moderator 对论文进行初步筛选，并采取一些措施来阻止试图强行绕过上诉过程而再次提交已经被拒稿的论文作者。对于进一步审核，则由平台董事会学科专业委员会推荐的领域专家志愿者完成。此外，调研中大部分预印本平台支持读者对已发布内容的开放评议，有助于进一步识别论文的学术性、净化预印本平台内容，同时也有利于作者对论文质量的提升。

以下重点对预印本服务质量控制措施进行介绍：

（1）审核论文上传者和作者的研究身份

论文上传者和作者的研究身份审核是预印本内容质量控制的第一步，审核控制点主要包括：注册邮箱、机构主页和信息、基金资助情况、同领域研究中的被认可度等。在六家平台中，有 3 家需要对用户注册和上传论文的资格进行审核，包括 arXiv、MDPI Preprint。arXiv 是当前全球规模最大的预印本平台，它对上传论文者的可信赖度进行严格审核，包括研究人员必须使用机构邮箱注册、主动声明科研身份、提交证明，推荐 arXiv 与 ORCID 的关联，审核通过的用户才可提交论文[6]。系统将确认作者的联系方式、并进行邮箱验证等；一旦发现伪造机构，将会被立即且永久地禁止提交。注册成功后，上传者还必须获得一位同领域科研人员的认可，才能顺利上传论文，通常这位研究人员需要在 arXiv 有一定的发文量。

（2）审核论文格式完整性

格式完整性审核的控制点包括：论文基本要素是否健全，格式是否符合平台需要格式。

除了 SSRN、ChemRxiv 调研暂未获得论文形式审核信息外，其余预印本服务都对论文形式方面进行质量控制。例如，arXiv 要求论文符合一般学术要求、基本要素完整，不接收只有摘要的论文，不接受不带引文的论文、图书或者评论、不包含原创、独立研究的报告；bioRxiv 要求上传文章须具备论文基本要素，格式为 PDF 或一般的文本格式。

（3）审核论文内容学术性和真实性

内容学术性和真实性审核，主要指预印本平台对论文学术性、合法性、真实性等方面的控制，不针对研究内容本身质量进行评价。审核控制点包括：学术性、非剽窃、不带有攻击性、不违反法律、无恶意攻击、诽谤性、涉密等不良信息和非法内容、无版权纠纷。以 arXiv 为例，其平台政策中明确说明需要“完全学术性（exclusively scientific research content）”的内容，并强调要与 18 个学科相关联。arXiv 不接收带有如下特征的论文：①包含有煽动性、虚假内容；②主题不合适，或有不如实、歪曲的表达；③论文标题、摘要导言运用具有特别戏剧性；④用户在短期内提交的论文内容重复；⑤版权纠纷作品，上传到 arXiv 的内容要求作者必须拥有完全的版权。此外，bioRxiv、SSRN、ChemRxiv 等都有类似的质量要求。

表 1 国内外知名预印本平台的质量控制内容

上传者和作者研究身份审核
格式完整性审核
内容学术性和真实性审核

arXiv[7]
邮箱后缀、所属机构、学科领域中研究学者的认可
论文符合一般学术要求、基本要素完整
学术性、学科关联性；学科领域专家参与论文评审
bioRxiv[8]
未明确说明
具有论文基本要素；必须为 PDF 或一般文本格式
非剽窃；不得带有攻击性、或者违背科学性的内容；未发表
SSRN[9]
未明确说明
未明确说明
文章内容符合学术性；
使用学术术语
PeerJ Preprint[10]
未明确说明
论文类型、字数限制；不同研究方向的论文规定了必备要素

得发表于其它预印本平台
ChemRxiv[11]
未明确说明
未明确说明
非剽窃；不得带有攻击性、或者违背科学性的内容；化学博士参与论文评审
MDPI Preprints[12]
上传论文需要先注册账户，提交前平台确认作者信息
不接受社论、或课程作业；具备基本论文要素
所有论文接受 MDPI 内部审核

从实践效果看，当前全球各领域预印本平台纷纷出现，论文数量不断增长，上述质量控制措施为预印本服务的有序健康发展提供了有效保障，是当前国际预印本服务的通行做

法。

3 ChinaXiv 质量控制机制

中国科学院科技论文预发布平台 ChinaXiv（http://chinaxiv.org）由中国科学院于 2016 年推出，面向全国科研人员，提供自然科学领域的中英文预印本服务。为保障平台内容的学术性、严谨性、权威性，ChinaXiv 按国际通行做法建立了质量控制机制，包括论文常规审核制、领域专家评审制、论文开放评议制、问题论文撤稿制和非诚信作者名单制等。

（1）论文常规审核制

ChinaXiv 设立质量控制审组负责每日论文预印本的审核发布；审核分为初审、终审两级，目的是最大限度保证论文审核的公平公正。同时，ChinaXiv 制定了《中国科学院科技论文预发布论文审核工作指南》（简称“审核工作指南”）和《中国科学院科技论文预发布平台论文准入规范》（简称“论文准入规范”），以保证论文审核工作有序、高效进行其中，审核工作指南主要对论文审核工作的流程和操作进行规范；准入规范，有效帮助审核人员判断哪些论文可以发布。

论文常规审核工作，主要包括以下三个方面：

①识别论文提交者和作者的研究身份。

论文准入规范明确要求上传者和作者具备正式的研究身份。ChinaXiv 建议用户使用机构邮箱注册、并进行邮箱和手机验证[13]。注册时建议作者可以选择提交 ORCID、机构主页、学术主页、简要学历、研究经历，以及在国内外同行评议期刊上发表过的论文等材料。一旦发现论文上传者或者作者研究机构或身份造假，或者未在正式刊物上发表过同主题的同行评审论文，那么论文将不能在 ChinaXiv 发布

②审核预发布论文的格式完整性、内容重复性、剽窃性、不涉及版权问题

ChinaXiv 系统能够自动检查论文题目、摘要、关键词、参考文献等必备要素的完备性，以及论文是否与平台已有论文重复、是否涉嫌剽窃。机器判断结果用以辅助人工审核。同时，上传至平台的论文，作者应保证不涉及版权纠纷。

③甄别预发布论文内容的研究性、科学性。

ChinaXiv 在网站中明确声明，作者向平台提交的科技论文应当具有科学性、创新性，具有一定学术水平和发布价值，坚决抵制违反学术道德和规范的学术不端行为，不抄袭、不剽窃、不造假，不涉及恶意攻击、诽谤性、涉密等不良信息和非法内容。同时，准入规范中明确提出，对于论文中有不当言论，或者是伪科学的论文不能在 ChinaXiv 平台发布。

通过上述三个方面的审核，ChinaXiv 平台的日常运营和内容质量得到有效保证。

（2）领域专家评审制

除了论文质量管理组，ChinaXiv 还充分发动科研人员力量参与预印本服务质量控制工作，包括 ChinaXiv 科学咨询委员会成员，以及中科院研究所具有副高级以上职称的科研人员等。当常规审核无法对论文的研究性、科学性做出判断时，则启用专家评审的流程，由

领域专家向 ChinaXiv 提供审核意见，然后由 ChinaXiv 平台做出论文是否发布的决定。

（3）论文开放评议制

ChinaXiv 遵循“先开放，后评议”原则，所有预印本发布后，将接受来自平台注册用户的评议。平台提供星级评分，和写评论包括公开评论、匿名评论、仅发给作者的功能。评议内容经审核团队审核后发布。对于开放评议差评多的论文，将会进入再审核，甚至撤稿流程。

（4）问题论文撤稿制

ChinaXiv 维护作者在平台上的论文发布权及被访问权，同时也要求论文作者遵守基本学术道德。平台论文一旦被发现有内容剽窃嫌疑、作者身份造假、研究内容有失科学性的情况，将提请平台负责人对论文内容进行审核，确定是否撤稿并发布相关申明。

（5）非诚信作者名单制

为杜绝非科技类、伪科学论文在平台发布，ChinaXiv 基于审核工作中发现的学术不诚信作者，如已被撤稿论文的作者等，建立非诚信作者名单，并由平台负责人核定。不诚信名单中的作者，其论文不再进入 ChinaXiv 审核与发布流程。

通过建立并实施上述机制，ChinaXiv 形成了以平台为主导，作者、用户、专家多方参与的预印本质量控制模式，实现了从论文上传、审核到发布的“全流程”质量控制 ChinaXiv 自 2016 年 6 月上线以来为科研人员提供了稳定预印本服务，累计接收并发布近 4000 篇。

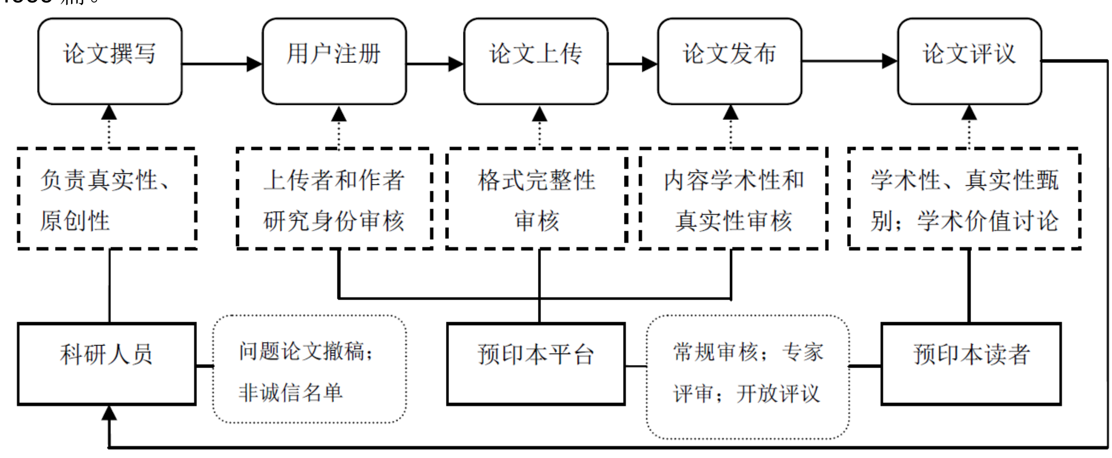


图 1 ChinaXiv 预印本质量控制机制示意图

4 ChinaXiv 质量控制机制的改进

从当前国际预印本的发展态势看，一方面，各领域预印本平台纷纷出现，论文数量不断上升，预印本在当前学术交流中扮演着越来越重要的角色；另一方面，各界对预印平台运营规范程度的要求进一步提高。NIH 在 2017 年 5 月发布的一份声明中明确提出，建议预印本平台确立透明公开的审核机制，对论文进行一定程度的质量控制，以保障科研人员对高质量学术内容的需求。同时国外也有基金组织呼吁建立对现有分散的预印本资源进行集成，建立一个预印本的中心资源（central service），从政策、技术、元数据、内容等方面

进行整体规划，提供核心服务。在这样的背景和趋势下，ChinaXiv 质量控制机制可以在如下方面提升和改进，从而更好地为科研人员提供持续、稳定、高质量的预印本服务。

首先，做好质量控制机制的公开透明。在国际预印本服务的发展过程中，平台运营方经常受到来自作者的质疑，认为预印本服务存在偏见。例如，日内瓦大学的一位量子物理学家怀疑 arXiv 存在“拉黑”作者的情况，原因是他的两名学生的一篇关于黑洞文章被 arXiv 拒稿。他认为这篇论文虽不够严谨，但作为预发布论文来说是合适的 [14]。类似的情况如不少人指责 arXiv 将一些论文（out there paper）放到一些活跃度和关注度不高分类下，阻碍了学术交流。虽然，这些质疑的合理性有待商榷，但反映出的问题值得重视。ChinaXiv 平台在未来发展过程中，应重视平台各类政策的公开和易于查找；同时加强对于审核人员的培训，做到规范操作，同时做好操作过程的记录和存档。

其次，充分发挥开放评议机制在质量控制中的作用。“先开放，后评议”是预印本作为新型学术交流方式的特点和优势，但从目前 ChinaXiv 的发展现状看，这种优势体现得还不够明显，开放评议机制在未来质量控制机制提升过程中可以发挥更重要的作用。ChinaXiv 应将用户对论文的评级、评论内容，在流程上与日常审核工作、领域专家审核有效对接。同时，参考社交网站对于评论发言政策，如问答系统 Quora，规范预印平台的开放评议行为，加强对评论内容的分析，提高论文审核的效率。

最后，ChinaXiv 应密切关注国际预印本平台发展动态，加强与其它预印本服务平台在质量控制政策、方法上的交流、研讨，进一步完善审核规范，促进简化审核流程、提高审核效率和准确度，防止伪科学论文混入平台，保障 ChinaXiv 的可持续健康发展。

参考文献

[1] 徐佳宁.国际电子预印本发展现状实证研究.图书馆建设,2010(10):42-46.
[2] 习惯 arXiv 的今天，我们的论文应该引用预印本吗？.[2017-08-15].https://baijiahao.baidu.com/s?id=1567623753629201&wfr=spider&for=pc
[3] NIH. Reporting Preprints and Other Interim ResearchProducts.[2017-08-15].https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-050.html
[4] WG Town, BA Vickery,JKuras,JR Weeks. Chemical e-journals, chemical e-preprints.Online Information Review , 2002, 26 (3):164-171.
[5] Wager E, aKleinert S. Responsible research publication: international standards for authors.Singapore: Imperial College Press / World Scientific Publishing,2010.
[6] Fledgling site challenges arXiv server.[2017-09-16].http://physicsworld.com/cws/article/news/2009/jul/15/fledgling-site-challenges-arxiv-server
[7] arXiv.https://arxiv.org/
[8] SM Awais.Principles of Transparency and Best Practice in Scholarly Publishing.[2017-09-16].http://www.trjemergmed.com/files/principles-of-transparency-and-best-practice-in-scholarly-publishing.pdf
[9] SSRN.https://www.ssrn.com/en/index.cfm/ssrn-faq/#paper_eligibility
[10] PeerJ Preprint.https://peerj.com/preprints-search/
[11] ChemRxiv. https://chemrxiv.org/
[12] MDPI Preprints.https://www.preprints.org/policy_and_process
[13] 夏莉霞,方卿.国外开放存取期刊的质量评价与控制研究综述.信息资源管理学报,2011(2):49-54.

[14] ZeeyaMerali.ArXiv rejections lead to spat over screening process.[2017-09-16].http://www.nature.com/news/arxiv-rejections-lead-to-spat-over-screening-process-1.19267